EM375 MECHANICAL ENGINEERING EXPERIMENTATION
POOLED VARIANCE

Many times, data are collected for a dependent variable, y, over a range of values for the independent variable, x. For example, the observation of fuel consumption might be studied as a function of engine speed while the engine load is held constant. If, in order to achieve a small variance in y, numerous repeated tests are required at each value of x, the expense of testing may become prohibitive. Reasonable estimates of variance can be determined by using the principle of *pooled variance* after repeating each test at a particular x only a few times.

Consider the following set of data for y obtained at various levels of the independent variable, x.

x	y
1	31, 30, 29
2	42, 41, 40, 39
3	31, 28
4	23, 22, 21, 19, 18
5	21, 20, 19, 18,17

The number of trials, mean, variance and standard deviation are presented in the next table.

x	n	y_{MEAN}	s_Y^2	S
1	3	30.0	1.00	1.00
2	4	40.5	1.67	1.29
3	2	29.5	4.50	2.12
4	5	20.6	4.30	2.07
5	5	19.0	2.50	1.58

These statistics represent the variance and standard deviation for each subset of data at the various levels of x. If we can assume that the same phenomena are generating random error at every level of x, the above data can be “pooled” to express a single estimate of variance and standard deviation. In a sense, this suggests finding a mean variance or standard deviation among the five results above. This mean variance is calculated by weighting the individual values with the size of the subset for each level of x. Thus, the POOLED VARIANCE is defined by:

$$S_P^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2 + \ldots + (n_k-1)S_k^2}{(n_1-1)+(n_2-1)+\ldots+(n_k-1)}$$

where n_1, n_2, . . . n_k are the sizes of the data subsets at each level of the variable x, and S_1^2, S_2^2, . . . S_k^2 are their respective variances.

The pooled variance of the data shown above is therefore:

$$S_P^2 = \frac{(3-1)\times 1.00 + (4-1)\times 1.67 + (2-1)\times 4.50 + (5-1)\times 4.30 + (5-1)\times 2.50}{(3-1)+(4-1)+(2-1)+(5-1)+(5-1)}$$

$$S_P^2 = 2.765$$

As an exercise, calculate the pooled variance and standard deviation for the following data:

x	y
1	29, 27
2	28, 25
3	33, 31.5
4	40, 39
5	41, 38
6	43, 36

Ans: $S_p^2 = 6.1875$